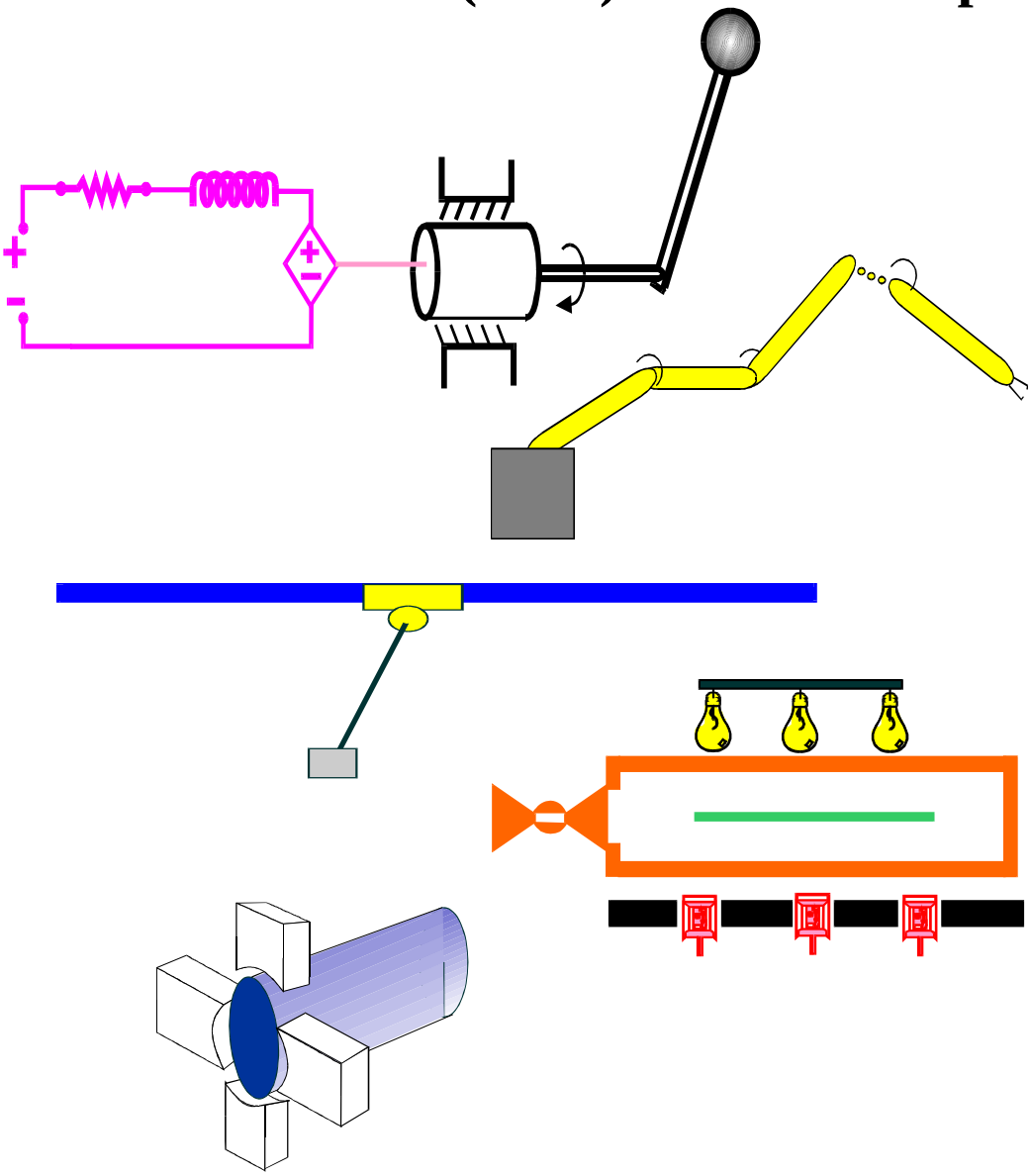
Clemson University
College of Engineering and Science
Control and Robotics (CRB) Technical Report



Number: CU/CRB/9/11/06/#1

Title: A Lyapunov-Based Method for Estimation of Euclidean Position of Static Features Using a Single Camera

Authors: V. K. Chitrakaran and D. M. Dawson

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
2006		**00-00-2006 to 00-00-2006**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
A Lyapunov-Based Method for Estimation of Euclidean Position of Static Features Using a Single Camera	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Clemson University,Department of Electrical & Computer Engineering,Clemson,SC,29634-0915**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**7**	

A Lyapunov-Based Method for Estimation of Euclidean Position of Static Features Using a Single Camera[1]

Vilas K. Chitrakaran† and Darren M. Dawson†

†Department of Electrical & Computer Engineering, Clemson University, Clemson, SC 29634-0915

E-mail: cvilas, ddawson@ces.clemson.edu

Abstract

In this paper, we present the design of an adaptive nonlinear algorithm for estimation of Euclidean position of features in a static environment in the field of view of a monocular camera. The development of the geometric model and camera motion kinematics is based on our previous work in [2]. This paper presents a new alternate approach to estimation of 3D coordinates of feature points that is simpler in mathematical formulation and easier to implement.

1 Introduction

The estimation of 3D Euclidean position of features in an environment from 2D images have applications in such diverse areas as autonomous vehicle navigation, visual servoing, 3D modeling, and geospatial mapping. In the computer vision community, this is typically recognized as the problem of identification of Structure from Motion (SFM), or Simultaneous Localization and Mapping (SLAM, [5]), where a vehicle such as a mobile robot is required to incrementally build a map of its environment as well as determine its position as it travels through the environment, scanning the scene using a vision system. Although the transformation of 3D Euclidean information by a perspective camera into 2D image space representation is inherently nonlinear, the most well known approaches are based on linearization based techniques such as extended Kalman filtering [5, 14]. Recently, the application of nonlinear system analysis tools towards a solution to this problem have begun to appear in works from many researchers [1, 8, 9], utilizing sliding mode and adaptive estimation techniques to build nonlinear observers.

The work presented in this paper is an evolution from our previous efforts [2] in the design and implementation of a nonlinear algorithm for 3D Euclidean position estimation. Similar to our previous approach, the camera motion is modeled in terms of the homography between two different views of the environment captured by the camera. One of the views is utilized as a constant image (corresponding to a reference position), while the other is a continuous temporal update of the 2D projection of the environment as seen by the camera while in motion. It is assumed that the camera is calibrated, and hence, the intrinsic calibration parameters are available. The kinematics of camera motion is formulated based on $2\frac{1}{2}$D visual servoing work presented in [13]. A stack of continuous motion estimators, described in detail in [3], is then utilized to estimate this kinematic signal in terms of every feature point. The result is a stack of camera velocity estimates that are scaled by the (constant) depth of the individual feature points from the camera relative to its reference location. The unknown scale factors are subsequently identified by using the adaptive least-squares technique [16], based on the satisfaction of a persistent excitation condition. With the unknown depths for every feature point identified, the 3D Euclidean position of all features are recovered. This is unlike our previous approach, where we developed a stack of Euclidean position estimators based on the image-space dynamics for every feature point. Hence, the mathematical formulation of the algorithm is simpler and more intuitive. Experimental results obtained from the same input data utilized in our previous work[2] seem to suggest that the new estimation algorithm converges faster, in a span of a few seconds as opposed to many tens of seconds.

This paper is organized as follows. A geometric model describing the evolution of image coordinates of various feature points in the environment in terms of camera motion is presented in Section 2. Section 3 presents the kinematics of the camera in terms of image space and partially reconstructed Euclidean information. The velocity estimation algorithm is described in Section 4, and is key to the development of the 3D Euclidean position estimation algorithm described in Section 5. Sections 6 and 7 demonstrate the performance of the position estimation algorithm through simulations and experimental results from our test-bed, respectively.

2 Vision System Geometric Model

Figure 1 shows the geometric relationship between a moving perspective camera and features on a static object in its field of view. The geometric model developed in this section is based on two views of the object from the camera; one of which is from a reference position denoted by $\mathcal{I}^*$, and the other is from a time varying current position denoted by $\mathcal{I}$. The origin of the camera frame is assumed to be coincident with the optical center of the camera. Let $x_f(t) \in \mathbb{R}^3$ represent the position of the object's body fixed frame

[1]This work was supported in part by two DOC Grants, an ARO Automotive Center Grant, a DOE Contract, a Honda Corporation Grant, and a DARPA Contract.

[2]A version of the work in [2] containing experimental results is currently under peer-review for consideration as a journal publication.

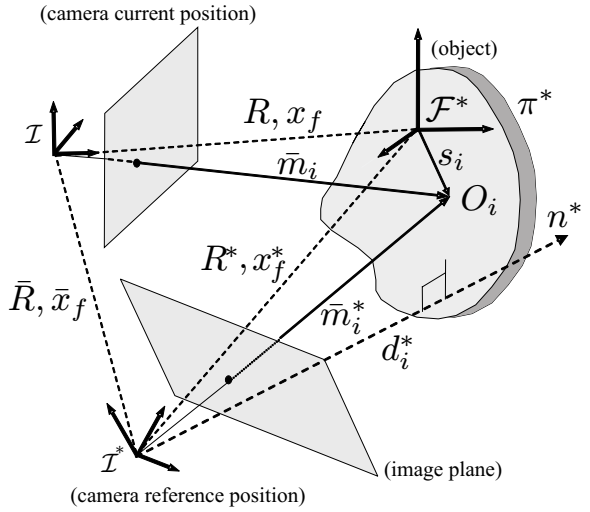


Figure 1: Geometrical relationships between the moving camera and the object.

$\mathcal{F}^*$ relative to the camera frame $\mathcal{I}$, and let $R(t) \in SO(3)$ represent the orientation between the object and the camera frame such that $R : \mathcal{F}^* \to \mathcal{I}$. The constant vector $s_i \in \mathbb{R}^3$ and the vector $\bar{m}_i(t) \in \mathbb{R}^3$ denote the 3D Euclidean position of the i^{th} feature point O_i relative to the object frame $\mathcal{F}^*$ and camera frame $\mathcal{I}$, respectively, and

$$\bar{m}_i \triangleq \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^T. \tag{1}$$

The constant terms $x_f^* \in \mathbb{R}^3$, $R^* \in SO(3)$ and $\bar{m}_i^* \in \mathbb{R}^3$ are similarly defined for the camera at the reference position denoted by $\mathcal{I}^*$. From the geometry between the coordinate frames depicted in Figure 1, it can be seen that

$$\bar{m}_i = x_f + Rs_i \tag{2}$$
$$\bar{m}_i^* = x_f^* + R^* s_i. \tag{3}$$

After eliminating the term s_i from (2) and (3), we have

$$\bar{m}_i = \bar{x}_f + \bar{R}\bar{m}_i^* \tag{4}$$

where $\bar{R}(t) \in SO(3)$ and $\bar{x}_f(t) \in R^3$ are defined as follows

$$\bar{R} = R(R^*)^T \qquad \bar{x}_f = x_f - \bar{R}x_f^*. \tag{5}$$

It is clear from (5) and Figure 1 that $\bar{R}(t)$ and $\bar{x}_f(t)$ describe the rotation and translation, respectively, of the camera frame $\mathcal{I}$ relative to the reference position $\mathcal{I}^*$.

Let three non-collinear feature points on the object define the plane π^* in Figure 1. If the constants $n^* \in \mathbb{R}^3$ and $d \in \mathbb{R}$ denote the normal vector and distance to this plane, respectively, from the camera reference position $\mathcal{I}^*$, then for all i feature points that lie on the plane, (4) can be expressed as follows

$$\bar{m}_i = \underbrace{\left(\bar{R} + \frac{\bar{x}_f}{d^*} n^{*T}\right)}_{H} \bar{m}_i^* \tag{6}$$

where $H(t) \in \mathbb{R}^{3\times 3}$ represents a Euclidean homography [6]. To facilitate the development of the relationship given in (6) in terms of image coordinates of the feature points as captured by the moving camera, we define the normalized Euclidean coordinates of the feature points, denoted by $m_i(t), m_i^* \in \mathbb{R}^3$ as

$$m_i \triangleq \frac{\bar{m}_i}{z_i} \qquad m_i^* \triangleq \frac{\bar{m}_i^*}{z_i^*}. \tag{7}$$

The corresponding projective pixel coordinates of the feature points are denoted by $p_i(t), p_i^* \in \mathbb{R}^3$ as follows

$$p_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T \qquad p_i^* = \begin{bmatrix} u_i^* & v_i^* & 1 \end{bmatrix}^T. \tag{8}$$

The image coordinates of the features and their normalized Euclidean coordinates are related by the pin-hole camera model [7] such that

$$p_i = Am_i \qquad p_i^* = Am_i^* \tag{9}$$

where $A \in \mathbb{R}^{3\times 3}$ is a known, constant, and invertible intrinsic camera calibration matrix. Utilizing (7) and (9), the relationship in (6) can be expressed in terms of image coordinates as follows

$$p_i = \underbrace{\frac{z_i^*}{z_i}}_{\alpha_i} \underbrace{A\left(\bar{R} + \bar{x}_h (n^*)^T\right) A^{-1}}_{G} p_i^* \tag{10}$$

where $\alpha_i(t) \in \mathbb{R}$ denotes the depth ratio, $\bar{x}_h(t) = \frac{\bar{x}_f}{d^*} \in \mathbb{R}^3$ denotes the scaled translation vector and $G(t) \in \mathbb{R}^{3\times 3}$ is a full rank homogeneous collineation matrix [12]. Given more than four pairs of correspondences $(p_i(t), p_i^*)$ coplanar with π^*, the set of linear equations in (10) can be solved for a unique $G(t)$ up to a scale factor by using techniques such as least squares minimization [12]. This approach is appropriate for visual servoing in structured environments, where we have the option of choosing feature points that are coplanar. In a structure-from-motion algorithm, we are interested in mapping the shape of objects in the environment around the camera, and feature points of interest tracked by the vision system will not lie on a plane. The key here is to utilize the Virtual Parallax algorithm, proposed in [13] and also described in detail in our previous work in [2], in order to compute $G(t)$ and $\alpha_i(t)$. Utilizing this method, any three feature points on the object can be chosen in order to define a 'virtual' plane π^*. The epipolar constraints governing the projection of the rest of the feature points on to this virtual plane are then utilized to develop an expression that facilitates the computation of $G(t)$. The Virtual Parallax algorithm requires at least eight feature points in order to compute the matrix $G(t)$, which can be subsequently used to uniquely determine $H(t)$ in (6), given that the intrinsic camera calibration matrix A is assumed to be known [12]. The matrix $H(t)$ can be decomposed to its constituent rotation matrix $\bar{R}(t)$ and scaled translation vector $\bar{x}_h(t)$ by utilizing the decomposition algorithms described in detail in works such as [7, 12].

Remark 1 *The reader is referred to [2] for computation of the homography matrix $H(t)$ and the depth ratios $\alpha_i(t)$ using the Virtual Parallax algorithm for non-coplanar feature points.*

3 Camera Kinematics

The translation and rotation kinematics of the camera frame $\mathcal{I}$ relative to the fixed position $\mathcal{I}^*$, denoted by $e_v(t) \in \mathbb{R}^3$ and $e_\omega(t) \in \mathbb{R}^3$, respectively, are developed in terms of the image coordinates of one of the feature points ($i = 1$ chosen here for the sake of simplicity) and the partially reconstructed Euclidean information from the decomposition of the homography matrix $H(t)$ as follows

$$e_v \triangleq \begin{bmatrix} u_1 - u_1^* & v_1 - v_1^* & -\ln(\alpha_1) \end{bmatrix}^T \quad (11)$$

$$e_\omega \triangleq u\phi \quad (12)$$

where $e_\omega(t)$ is expressed in terms of the axis-angle representation [17] of $\bar{R}(t)$ defined in (5) such that $u(t) \in \mathbb{R}^3$ represents a unit rotation axis, and $\phi(t) \in \mathbb{R}$ denotes the rotation angle about $u(t)$, assumed to be confined to the region $-\pi < \phi(t) < \pi$. After taking the time derivative of (11) and (12), the camera kinematics can be written in the following form

$$\dot{e} = Jv \quad (13)$$

where $e(t) = \begin{bmatrix} e_v^T(t) & e_\omega^T(t) \end{bmatrix}^T \in \mathbb{R}^6$ and $v(t) = \begin{bmatrix} v_c^T(t) & \omega_c^T(t) \end{bmatrix}^T \in \mathbb{R}^6$, where $v_c(t), \omega_c(t) \in \mathbb{R}^3$ denote the linear and angular velocities of the camera relative to the reference position $\mathcal{I}^*$ but expressed in the local frame $\mathcal{I}$. In (13), $J(t) \in \mathbb{R}^{6\times 6}$ is the following Jacobian-like matrix

$$J = \begin{bmatrix} \frac{\alpha_1}{z_1^*} A_{e1} & A_{e1}[m_1]_\times \\ 0_{3\times 3} & -L_\omega \end{bmatrix} \quad (14)$$

where $[m_i]_\times \in \mathbb{R}^{3\times 3}$ denotes the skew-symmetric form of $m_i(t)$, $O_{3\times 3} \in \mathbb{R}^{3\times 3}$ denotes a 3×3 zero matrix, $A_{ei}(t) \in \mathbb{R}^{3\times 3}$ is a function of the camera intrinsic calibration parameters and image coordinates of the i^{th} feature point as shown below

$$A_{ei} \triangleq A - \begin{bmatrix} 0 & 0 & u_i \\ 0 & 0 & v_i \\ 0 & 0 & 0 \end{bmatrix} \quad (15)$$

and $L_\omega(t) \in \mathbb{R}^{3\times 3}$ is defined as follows

$$L_\omega \triangleq I_3 - \frac{\phi}{2}[u]_\times + \left(1 - \frac{sinc(\phi)}{sinc^2\left(\frac{\phi}{2}\right)}\right)[u]_\times^2 \quad (16)$$

where $sinc(\phi(t)) \triangleq \dfrac{\sin\phi(t)}{\phi(t)}$.

4 Velocity Estimation

Unlike our previous approach [2], where we utilized the image-space dynamics for every feature point in order to develop a stack of estimators for their Euclidean positions, the estimation algorithm presented in this paper is based on a stack of velocity estimators; one per feature point. To facilitate this development, the kinematic expression in (13) is rewritten for every feature point tracked by the vision system as follows

$$\dot{e}_i = J_{ki} v_i \quad (17)$$

where

$$J_{ki} = \begin{bmatrix} \alpha_i A_{ei} & A_{ei}[m_i]_\times \\ 0_{3\times 3} & -L_\omega \end{bmatrix} \quad (18)$$

$$v_i = \begin{bmatrix} \frac{1}{z_i^*} v_c^T & \omega_c^T(t) \end{bmatrix}^T. \quad (19)$$

Note that the first three elements of $v_i(t) \in \mathbb{R}^6$ in (19) denote the linear velocity of the i^{th} feature point scaled by the constant depth z_i^*. The scaled motion signal $\dot{e}_i(t) \in \mathbb{R}^6$ is estimated utilizing the following continuous estimator

$$\dot{\hat{e}}_i \triangleq \int_{t_0}^{t} (K_i + I_{6\times 6})\tilde{e}_i(\tau)d\tau + \int_{t_0}^{t} \rho_i \text{sgn}(\tilde{e}_i(\tau))\, d\tau + (K_i + I_{6\times 6})\tilde{e}_i(t) \quad (20)$$

where $\dot{\hat{e}}_i(t) \triangleq \begin{bmatrix} \dot{\hat{e}}_{vi}^T(t) & \dot{\hat{e}}_\omega^T(t) \end{bmatrix}^T \in \mathbb{R}^6$ denotes the estimate of the signal $\dot{e}_i(t)$, $\tilde{e}_i(t) \triangleq e_i(t) - \hat{e}_i(t) \in \mathbb{R}^6$ is the estimation error in $e_i(t)$, $K_i, \rho_i \in \mathbb{R}^{6\times 6}$ are positive definite constant diagonal gain matrices, and $\text{sgn}(\tilde{e}_i)$ denotes the signum function applied to each element of the vector $\tilde{e}_i(t)$. The development of the above estimator is described in detail in [3]. To summarize, it was shown that the estimator in (20) asymptotically identifies the signal $\dot{e}_i(t)$ (*i.e.*, $\dot{\hat{e}}_i(t) \to \dot{e}_i(t)$ as $t \to \infty$), provided that the j^{th} diagonal element of the gain matrix ρ_i and the j^{th} element of the vectors $\dddot{e}_i(t)$ and $\ddddot{e}_i(t)$ satisfies the following condition for all i feature points

$$[\rho_i]_j \geq \left|[\dddot{e}_i]_j\right| + \left|[\ddddot{e}_i]_j\right| \quad \forall j = 1, 2, ...6. \quad (21)$$

The analysis presented in [3] demonstrated that the estimator in (20) satisfies the following property

$$\tilde{e}_i(t), \dot{\tilde{e}}_i(t) \in \mathcal{L}_\infty \cap \mathcal{L}_2. \quad (22)$$

The result in (22) is critical for the subsequent Euclidean position estimator design. Since $J_{ki}(t)$ is full rank, and composed of known signals, the estimate for the scaled velocity signal, denoted by $\hat{v}_i(t) = \begin{bmatrix} \frac{1}{z_i^*}\hat{v}_c^T & \hat{\omega}_c^T(t) \end{bmatrix}^T \in \mathbb{R}^6$ can be obtained from (17) and (20) as follows

$$\hat{v}_i = J_{ki}^{-1}\, \dot{\hat{e}}_i \quad (23)$$

and as a consequence of the result in (22), it can be shown [2] that the estimation error in scaled velocity, denoted by $\tilde{v}_i(t) \triangleq v_i(t) - \hat{v}_i(t) \in \mathbb{R}^6$ satisfies the following

$$\tilde{v}(t) \in \mathcal{L}_\infty \cap \mathcal{L}_2 \quad (24)$$

$$\text{and } \hat{v}_i(t) \to v_i(t) \text{ as } t \to \infty. \quad (25)$$

5 Euclidean Position Estimation

From (17), the kinematic signal $\dot{e}_i(t)$ for the i^{th} feature point can be written in terms of the scaled velocity of any other j^{th} feature point, $j \neq i$. Choosing $j = 1$ for the sake of simplicity, it can be shown that

$$\dot{e}_i = J_{ki}\text{diag}(\lambda_i, \lambda_i, \lambda_i, 1, 1, 1)\, J_{k1}^{-1}\dot{e}_1 \quad (26)$$

where

$$\lambda_i = \frac{z_1^*}{z_i^*} \quad (27)$$

is the unknown depth ratio, and $\text{diag}(\cdot) \in \mathbb{R}^{6\times 6}$ denotes a diagonal matrix. In terms of the estimates $\dot{\hat{e}}_i(t)$, the expression in (26) can be rewritten in the following manner

$$\dot{\bar{e}}_i = J_{ki}\text{diag}\left(\lambda_i, \lambda_i, \lambda_i, 1, 1, 1\right) J_{k1}^{-1}\dot{\hat{e}}_1 + \eta_i \tag{28}$$

where $\eta_i(t) \in \mathbb{R}^6$ is defined as follows

$$\eta_i = J_{ki}\text{diag}\left(\lambda_i, \lambda_i, \lambda_i, 1, 1, 1\right) J_{k1}^{-1}\dot{\bar{e}}_1 - \dot{\bar{e}}_i. \tag{29}$$

After substituting for $J_{ki}(t)$ in (28) and simplifying the resulting expression, the translational kinematics for the i^{th} feature point can be expressed as

$$\bar{y}_i = \lambda_i \bar{h}_{1i} + \bar{h}_{2i} + \eta_{vi} \tag{30}$$

where $\bar{y}_i(t) \triangleq \dot{\bar{e}}_{vi}(t) \in \mathbb{R}^3$, $\eta_{vi}(t) \in \mathbb{R}^3$ is composed of the first three elements of $\eta_i(t)$, and $\bar{h}_{1i}(t), \bar{h}_{2i}(t) \in \mathbb{R}^3$ are the following measurable signals

$$\bar{h}_{1i} = \frac{\alpha_i}{\alpha_1}\left(A_{ei}A_{e1}^{-1}\dot{\hat{e}}_{v1} + A_{ei}\left[m_1\right]_\times L_\omega^{-1}\dot{\hat{e}}_w\right) \tag{31}$$

$$\bar{h}_{2i} = -A_{ei}\left[m_i\right]_\times L_\omega^{-1}\dot{\hat{e}}_w. \tag{32}$$

In terms of a subsequently designed estimate $\hat{\lambda}_i(t) \in \mathbb{R}$ for the depth ratio λ_i, the estimate for $\bar{y}_i(t)$, denoted by $\hat{\bar{y}}_i(t) \in \mathbb{R}^3$, can be expressed as follows

$$\hat{\bar{y}}_i = \hat{\lambda}_i \bar{h}_{1i} + \bar{h}_{2i}. \tag{33}$$

After denoting $\tilde{\bar{y}}_i(t) \triangleq \bar{y}_i(t) - \hat{\bar{y}}_i(t) \in \mathbb{R}^3$ as the estimation error signal, it can be seen from (30) and (33) that

$$\tilde{\bar{y}}_i = \tilde{\lambda}_i \bar{h}_{1i} + \eta_{vi} \tag{34}$$

where $\tilde{\lambda}_i(t) \triangleq \lambda_i(t) - \hat{\lambda}_i(t) \in \mathbb{R}$. Based on the subsequent Lyapunov-based stability analysis, the depth ratio estimate $\hat{\lambda}_i(t)$ is updated per the following adaptive least-squares update law

$$\dot{\hat{\lambda}}_i = \beta_i L_i \bar{h}_{1i}^T \tilde{\bar{y}}_i \tag{35}$$

where $\beta_i > 1 \in \mathbb{R}$ is a positive constant, $L_i(t) \in \mathbb{R}$ is an estimation gain that is computed as follows

$$\frac{d}{dt}L_i^{-1} = \bar{h}_{1i}^T \bar{h}_{1i} \tag{36}$$

and initialized such that $L_i^{-1}(0) > 0$.

5.1 Stability Analysis

Theorem 1 *The update law defined in (35) ensures that $\hat{\lambda}_i(t) \to \lambda_i$ as $t \to \infty$ provided that the following persistent excitation condition is satisfied*

$$\gamma_1 \leq \int_{t_0}^{t_0+T} \bar{h}_{1i}^T(\tau)\bar{h}_{1i}(\tau)d\tau \leq \gamma_2 \tag{37}$$

where $t_0, T, \gamma_1, \gamma_2 \in \mathbb{R}$ are positive constants.

Proof: Let $V(t) \in \mathbb{R}$ denote a non-negative scalar function defined as follows

$$V \triangleq \frac{1}{2}\tilde{\lambda}_i L_i^{-1}\tilde{\lambda}_i. \tag{38}$$

After taking the time derivative of (38) and substituting the adaptive update law in (35), the following expression can be obtained

$$\begin{aligned}\dot{V} &= -\beta_i \tilde{\lambda}_i L_i^{-1} L_i \bar{h}_{1i}^T\left(\tilde{\lambda}_i \bar{h}_{1i} + \eta_{vi}\right) \\ &\leq -\beta_i\left|\tilde{\lambda}_i\right|^2\left\|\bar{h}_{1i}\right\|^2 + \left|\beta_i \tilde{\lambda}_i \bar{h}_{1i}^T \eta_{vi}\right|.\end{aligned} \tag{39}$$

After utilizing the nonlinear damping argument [10], the second term in (39) can be upper-bounded as follows

$$\left|\beta_i \tilde{\lambda}_i \bar{h}_{1i}^T \eta_{vi}\right| \leq \left|\tilde{\lambda}_i\right|^2\left\|\bar{h}_{1i}\right\|^2 + \beta_i^2\left\|\eta_{vi}\right\|^2, \tag{40}$$

hence, the time derivative of (38) can be upper-bounded in the following manner

$$\dot{V} \leq -\mu_i\left|\tilde{\lambda}_i\right|^2\left\|\bar{h}_{1i}\right\|^2 + \beta_i^2\left\|\eta_{vi}\right\|^2 \tag{41}$$

where $\mu_i = (\beta_i - 1) \in \mathbb{R}$ is a scalar positive constant. From (22), (24) and (29), we can show that $\eta_{vi}(t) \in \mathcal{L}_2 \cap \mathcal{L}_\infty$, and hence,

$$\int_{t_0}^{t} \mu_i\left|\tilde{\lambda}_i(\tau)\right|^2\left\|\bar{h}_{1i}(\tau)\right\|^2 d\tau \leq V(0) - V(\infty) + \varepsilon_i \tag{42}$$

where $\varepsilon_i \in \mathbb{R}$ is a positive constant such that

$$\int_{t_0}^{t} \beta_i^2\left\|\eta_{vi}(\tau)\right\|^2 d\tau \leq \varepsilon_i. \tag{43}$$

From (42), it can be concluded that $\tilde{\lambda}_i(t)\bar{h}_{1i}(t) \in \mathcal{L}_2$. As evident from (38) and (42), $V(t) \leq V(0) + \varepsilon$ for any time t, and hence, $V(t) \in \mathcal{L}_\infty$; therefore, $\tilde{\lambda}_i(t) \in \mathcal{L}_\infty$. Since the term $\bar{h}_{1i}(t)$ defined in (31) is composed of bounded signals, $\tilde{\lambda}_i(t)\bar{h}_{1i}(t) \in \mathcal{L}_\infty \cap \mathcal{L}_2$. From (34), $\tilde{\bar{y}}_i(t) \in \mathcal{L}_\infty$, and it follows from (35) that $\dot{\hat{\lambda}}_i(t), \dot{\tilde{\lambda}}_i(t) \in \mathcal{L}_\infty$. Following the analysis presented in [3], we can show that the signal $\dot{\tilde{\bar{y}}}_i(t) \in \mathcal{L}_\infty$, and hence, from taking the time derivative of (31), it can be shown that $\dot{\bar{h}}_{1i}(t) \in \mathcal{L}_\infty$. Hence, we have established that $\tilde{\lambda}_i(t)\bar{h}_{1i}(t)$ is uniformly continuous, based on the fact that $\tilde{\lambda}_i(t)\bar{h}_{1i}(t)$ and $\frac{d}{dt}\left(\tilde{\lambda}_i(t)\bar{h}_{1i}(t)\right)$ are bounded signals [4]. Utilizing Barbalat's lemma [4], it can be concluded that

$$\tilde{\lambda}_i(t)\bar{h}_{1i}(t) \to 0 \text{ as } t \to \infty. \tag{44}$$

If the signal $\bar{h}_{1i}(t)$ satisfies the persistent excitation condition [16] given in (37), then it can be concluded from (44) that

$$\tilde{\lambda}_i(t) \to 0 \text{ as } t \to \infty. \tag{45}$$

□

Remark 2 *From (31), it can be observed that the persistent excitation condition in (37) is easily satisfied if the camera has non-zero translational velocity.*

Remark 3 *$\hat{\lambda}_i(t)$ provides an estimate for $\frac{z_1^*}{z_i^*}$ (see (27)). If z_1^* is known a priori from secondary measurements, an estimate for the 3D Euclidean coordinates of the i^{th} feature*

point at the reference position, denoted by $\hat{\bar{m}}_i^*(t) \in \mathbb{R}^3$, can be obtained as follows

$$\hat{\bar{m}}_i^*(t) = \frac{z_1^*}{\hat{\lambda}_i(t)} A^{-1} p_i^*. \tag{46}$$

If the Euclidean distance between any two features i and j ($i \neq j$) is known (i.e., $\|\bar{m}_i^* - \bar{m}_j^*\|$), then from (35), (46) and the above theorem, it can be seen that

$$\lim_{t\to\infty} \left\| \frac{1}{\hat{\lambda}_i(t)} A^{-1} p_i^* - \frac{1}{\hat{\lambda}_j(t)} A^{-1} p_j^* \right\| = \frac{1}{z_1^*} \|\bar{m}_i^* - \bar{m}_j^*\|. \tag{47}$$

The scale factor z_1^* can be recovered from the above expression and utilized in (46) in order to determine the Euclidean coordinates.

6 Simulation Results

For validation of the position estimation algorithm presented in the previous section, we simulated a static object with 12 feature points in the field of view of a moving camera. The trajectory for camera translational velocity was chosen as follows

$$v_c(t) = [\cos(t), -\sin(t), 0.1\sin(t)]^T \text{ (m/s)}. \tag{48}$$

For the sake of simplicity, the camera intrinsic calibration matrix in (9) was assumed to be the identity matrix. The image coordinates of the feature points on the static object were updated utilizing the kinematics presented in Section 3. The estimator gains were chosen through trial and error for all i feature points as follows

$$\begin{aligned} K_i &= \text{diag}(10, 10, 10, 10, 10, 10), \\ \rho_i &= \text{diag}(1, 1, 1, 1, 1, 1), \\ \beta_i &= 1.5. \end{aligned} \tag{49}$$

Figure 2 shows the convergence of the inverse depth estimates $\hat{\lambda}_i(t)$ for all feature points on the object. The estimation errors are shown in Figure 3.

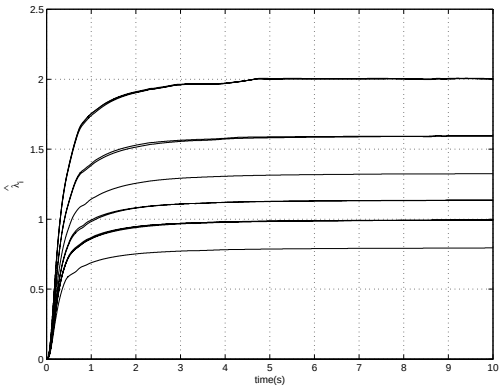


Figure 2: Convergence of scaled inverse depth estimates (dimensionless) for all feature points on the object (simulated camera motion).

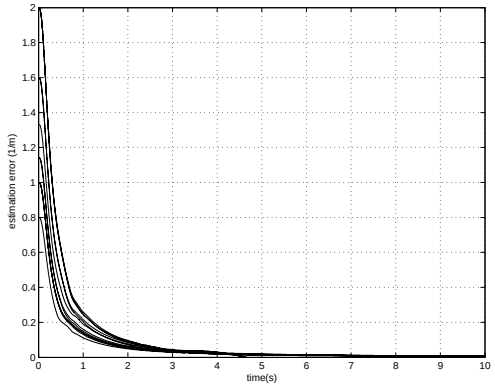


Figure 3: Error in depth estimation (simulated camera motion).

7 Experimental Results

Figure 4 shows a single frame from a video sequence utilized to verify the practical performance of the estimator. The video was captured from a monocular gray-scale camera (Sony XC-ST50) found to have the following intrinsic calibration parameters

$$A = \begin{bmatrix} 797.5 & 0 & 312.9 \\ 0 & 819.4 & 231.8 \\ 0 & 0 & 1 \end{bmatrix}. \tag{50}$$

The camera was mounted on the end-effector of a Puma 560 industrial manipulator arm and programmed to move along a smooth closed trajectory. The camera was interfaced to our vision system through an Imagenation PXC200-AF framegrabber, and triggered to capture images at the rate of 20 frames per second through an external time source. We utilized an implementation of the Lucas-Kanade feature tracking algorithm [11] provided in the OpenCV computer vision library [15] to detect and track feature points in the video stream. The implementation of feature tracking and depth estimation programs were in C++. The estimator utilized the gains shown in (49). Table 1 shows a comparison between the actual and the estimated lengths in the scene. The time evolution of the inverse depth estimates are shown in Figure 5. Note that the depth estimates converge quickly, in a matter of a few seconds as opposed to several tens of seconds required by the estimator design based on pixel dynamics in our previous work [2].

8 Conclusions

This paper presented an algorithm to identify the Euclidean coordinates of features points in a static environment using a calibrated monocular moving camera. The work presented here utilizes an approach to geometric modeling similar to our previous work [2], but describes the development of an alternate nonlinear estimation algorithm which the authors believe to be simpler in mathematical formula-

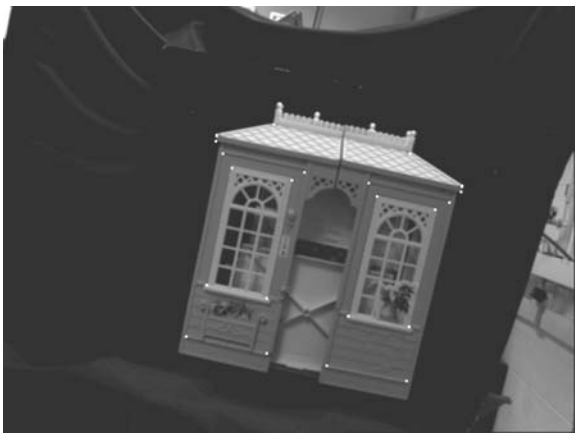


Figure 4: A frame from the doll-house video sequence used in the experimental validation. The white dots are the tracked feature points.

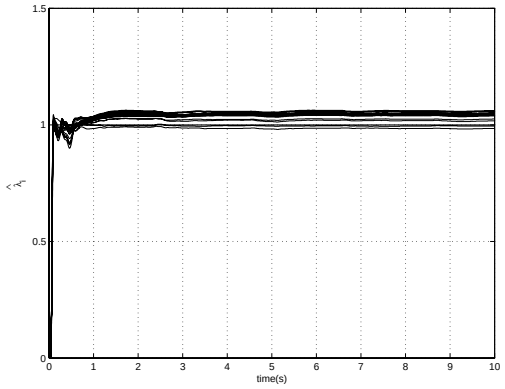


Figure 5: The time evolution of inverse depth estimates from the experimental test-bed.

Object	Act. dim. (cm.)	Est. dim. (cm.)
Length I	23.6	23.4
Length II	39.7	40.4
Length III	1.0	1.0
Length IV	13.0	13.1
Length V	10.0	10.1
Length VI	19.8	19.7
Length VII	30.3	30.1

Table 1: Estimated dimensions from the experimental test-bed.

tion, and more intuitive. The performance of the algorithm has been verified through simulations and an experimental implementation.

References

[1] X. Chen, and H. Kano, "State Observer for a Class of Nonlinear Systems and Its Application to Machine Vision," *IEEE Transactions on Automatic Control*, Vol. 49, No. 11, 2004.

[2] V. K. Chitrakaran, D. M. Dawson, J. Chen, and W. E. Dixon, "Euclidean Position Estimation of Features on a Moving Object Using a Single Camera: A Lyapunov-Based Approach," *Proc. of the American Control Conference*, pp. 4601-4606, June 2005.

[3] V. Chitrakaran, D. M. Dawson, W. E. Dixon, and J. Chen, "Identification of a Moving Object's Velocity with a Fixed Camera," *Automatica*, Vol. 41, No. 3, pp. 553-562, March 2005.

[4] W. E. Dixon, A. Behal, D. M. Dawson, and S. Nagarkatti, *Nonlinear Control of Engineering Systems: A Lyapunov-Based Approach*, Birkhäuser Boston, 2003.

[5] H. Durrant-Whyte, and T. Bailey, "Simultaneous Localization and Mapping: Part I," *IEEE Robotics and Automation Magazine*, Vol. 13, No. 2, pp. 99-108, June 2006.

[6] O. Faugeras and F. Lustman, "Motion and Structure From Motion in a Piecewise Planar Environment," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 2, No. 3, pp. 485-508, 1988.

[7] O. Faugeras, *Three-Dimensional Computer Vision*, The MIT Press, Cambridge, Massachusetts, 2001.

[8] X. Hu, and T. Ersson, "Active State Estimation of Nonlinear Systems," *Automatica*, Vol. 40, pp. 2075-2082, 2004.

[9] M. Jankovic, and B. K. Ghosh, "Visually Guided Ranging from Observations of Points, Lines and Curves via an Identifier Based Nonlinear Observer," *Systems and Control Letters*, Vol. 25, pp. 63-73, 1995.

[10] M. Krstić, I. Kanellakopoulos, and P. Kokotović, *Nonlinear and Adaptive Control Design*, New York, NY: John Wiley and Sons, 1995.

[11] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *International Joint Conference on Artificial Intelligence*, pp. 674-679, 1981.

[12] Y. Ma, S. Soatto, J. Košecká, and S. Sastry, *An Invitation to 3D Vision*, Springer-Verlag, ISBN: 0387008934, 2003.

[13] E. Malis, and F. Chaumette, "2 1/2 D Visual Servoing with Respect to Unknown Objects Through a New Estimation Scheme of Camera Displacement," *International Journal of Computer Vision*, Vol. 37, No. 1, pp. 79-97, 2000.

[14] J. Oliensis, "A Critique of Structure From Motion Algorithms," *Computer Vision and Image Understanding*, Vol. 80, No. 2, pp. 172-214, 2000.

[15] Open Source Computer Vision Library, http://www.intel.com/technology/computing/opencv/index.htm.

[16] S. Sastry, and M. Bodson, *Adaptive Control: Stability, Convergence, and Robustness*, Prentice Hall, Inc: Englewood Cliffs, NJ, 1989.

[17] M. W. Spong and M. Vidyasagar, *Robot Dynamics and Control*, John Wiley and Sons, Inc: New York, NY, 1989.